Reviewer 2

R1:The knowledge on shoot to root relationships in grazed pastures is limited and this hinders not only the ability to model the grazing systems for their potential to accumulate carbon in soil but also to develop superior cultivars that withstand different intensities of grazing. The authors conducted an experiment in a common garden setting under low soil fertility conditions to evaluate the independent and combined effects of defoliation intensity and frequency, and cultivar on root production of a widely utilized pasture grass species of the southeastern United States, *Paspalum notatum* Flüegge (bahiagrass)*.* The two main objectives of the study were: (i) to isolate the effects of defoliation intensity, frequency and cultivar on belowground production, and (ii) evaluate the relationship between aboveground and belowground growth. The authors tested four specific hypotheses to evaluate shoot to root relationships of four cultivars (two with decumbent and two with upright growth habit). They found that severe defoliation (to 5 cm) improved shoot production for all four different cultivars, while it reduced root production across cultivars, regardless of defoliation frequency. They found no significant relationship between shoot and root production. Since the results from this study showed weak coupling between aboveground and belowground productivity, authors cautioned  against use of simple aboveground proxies to predict variations in root production in grasslands. In addition, the authors also pointed out the importance of cultivar level differences in response to defoliation management which need to be considered in modeling grassland management as well as developing superior cultivars through selection and/or breeding.

I find this as an important area of grassland research with broader application for modeling, agronomy and biology of grassland species. The research reported is important and it advances the current knowledge on shoot to root relationships of bahiagrass cultivars. The paper is well-written and the authors have done a very good job in conducting the experiment, and analyzing and interpreting the results. The authors are fully aware of the limitations of their study and these were stated clearly in the Discussion section. The paper is suitable for publication in Tropical Grasslands – Forrajes Tropicales after revision. I have the following suggestions for further improvement.

*We thank reviewer 2 for a positive, constructive review of our work.*

R2: Lines 150-152: The objectives stated here in the Introduction section may be used to structure the Results and Discussion sections. Results are presented on shoot model first followed by root model and then root to shoot relationships. It may be better to revise objectives to fit to the current structure or modify the current structure according to the objectives.

*We thank the reviewer for this helpful suggestion. We have reorganized our discussion section to follow the logic and ordering of our hypotheses, and grouped according to three main themes of: 1. Impacts of severe defoliation, 2. Cultivar variability, and 3. Spatial correlations in root and shoot growth. We carefully considered the results section, but have decided to leave the prior structure in place. Ultimately, the flow would be extremely confusing compared to the structure of our data. However, we think that the discussion section is a great place to highlight the connections and bring everything from introduction full circle, and appreciate the opportunity to make this improvement.*

R3: Lines 191: How about using SI units in Table S1?

*Corrected.*

R4: Lines 195-196: How about waterlogging effect on root development? Are there differences among the four cultivars tested? Need to include this information, if available.

*Unfortunately, we do not have any additional data on waterlogging stress as it impacts these cultivars. We note that this is a very typical season condition for a large amount of bahigrass production in the southern portion of the state. It is safe to assume that all the cultivars were equally exposed to this stress, so the root production results reported in this study in fact already bake in the impacts of waterlogging, whatever they may be.*

R5: Line 382: The Discussion section may be restructured based on revised objectives with clear sub-headings.

*We appreciate this comment, and we have restructured the discussion accordingly.*

Line 489: This section may be condensed to state a few clear messages from the study.

Line 542; 565: Check the complete title.

Responses to line-comments from editor

Line 155: Would above-ground production increase? (Lyle)

*Good point. We have clarified our hypothesis 2 in terms of shoot production. The text now reads, “Shoot production would either remain flat or decrease somewhat, as the higher level of stress over-rides compensatory growth mechanisms”.*

Line 185: Please use SI units in Supplement S1.

*Corrected.*

Line 187-188: Are there differences between the 4 cultivars tested in their level of tolerance to waterlogging? Reviewer

*As we discuss above, we do not have specific information on tolerance to waterlogging for our cultivars. We note on lines 189-191 that the waterlogging was observed in all plots, with no evident variation or relief in our experimental area. Thus, we feel it is safe to assume that all of our plants were equally impacted. Moreover, this condition is totally normal and widespread in Southern Florida during the summer growing season. We agree with the reviewer that this is a topic worth investigating more in-depth in future.*

Line 266: Please define (Lyle)

*We have expanded the text, and supplied a reference, so that it now reads: In the case of root allocation, we further analyzed all the pair-wise contrasts among cultivars (n = 6 contrasts), by taking the difference for each coefficient at each iteration of the Markov Chain Monte Carlo sampler, the computational algorithm by which Bayesian models are fit (Gelman et al. 2013)*

Line 276: Please define (Lyle)

*We have updated the text to define this acronym and supplied a new reference for the reader.*

Line 342-344: This cannot be correct.

*Thank you for pointing out this legend was confusing. We have corrected to clarify that the comparison is higher or lower, whereas before we had just said higher. The text now reads: “Where the entire 95% credible interval falls above or below zero, we can interpret that as a 97.5+% Bayesian probability of the first cultivar having a higher or lower root allocation, respectively, than the second cultivar.”*

Line 380: Was there any possibility that the excss water created a situation similar to insufficient water? (Lyle)

*Great question! We do not believe that our data are in a position to inform on this question, however it would stand to reason that there may be overlaps in some of the stress response mechanisms. One generally noted response to hypoxia during the wet summer growing period is a sloughing of deeper roots. The in-growth core method would need substantial revision to address this process, but it is certainly something worth considering for follow-up studies. The periodic root sloughing may be important mechanisms for root C input.*

Line 385: Is it intensity or severity? I interpret intensity as meaning quickly.

*Good catch. We realize we were using intensity and severity interchangeability, and have now standardized the manuscript to using the term “severity”.*

Line 406: Do you consider your study was long-term? (Lyle)

*Good question. No, we do not. We have reworded to emphasize that we are specifically contrasting our study from long-term studies.*

Line 465: What do you mean by managed? Most graziers would consider they manage their rangelands and they are far from monospecific. Do you mean areas sown with exotic species? (Lyle)

*Good point. We have reworded to focus on planted pastures, particularly in the subtropical climatic zones.*

Line 475: Why only ecologists? Why not agronomists and physiologists? Reviewer (Lyle?)

*Thank you for pointing out our omission. We have modified to be more inclusive.*

Line 483: Root turnover is also important for carbon sequestration. Reviewer

Is C storage not also relevant? (Lyle)

*We are unsure if this is meant as suggestion or comment. Will be happy to respond accordingly once clarified.*

Reviewer 1

R1: This is a disappointing paper. Its conclusions might be a useful indication of the interactions between defoliation frequency and severity on root growth in a range of *P. notatum* cultivars on waterlogged soils in southern Florida. But how relevant to normal pastures are root data of swards growing on waterlogged soil? I suspect not much.

*We are sorry that the reviewer was disappointed in the scope of our work. However, we disagree in terms of the generalizability of our study. We appreciated that reviewer 2, by contrast, noted our hard work to qualify our results. The climatic and edaphic conditions of Florida are unique in some ways, but in other ways generalize quite nicely to other subtropical pastures. For instance, the Brazilian Cerrado has many areas with high rainfall, coupled to low fertility, acidic soils. Perhaps more importantly, our main conclusions – surrounding our hypotheses 1-4, which we have more clearly structured our discussion around – relate to patterns of carbon allocation and presumed trade-offs in allocation evoked by defoliation treatments. We doubt that these results hinge critically on our unique growing environment. The reviewers point is well taken though, that much more work needs to be done across pasture types in various geographic locations.*

R2: The authors report data from a simple agronomic design of 4 replicates of 4 cultivars \* 2 cutting frequencies \* 2 cutting heights. The authors mention (Line 42) the need for “season-long” evaluations. The data they report are total harvested dry matter over just 16 weeks and a final measurement of root ingrowth at the end of this period. This is only a fraction of the growing season in sub-tropical Florida.

Given the short duration of the experiment in just one year, the data are of doubtful relevance to any practical cattle production system. Moreover, the authors provide no data of any sward characteristics such as LAI before and after the defoliation treatment that might give some insights into the effects of the imposed treatments.

*We respectfully disagree with the sweeping assertion that our data do not capture an adequate growing season, and thus are of doubtful relevance. Our shoot average production was 290 g/m2. This accords beautifully with published literature on annual dry biomass production in unfertilized Bahiagrass. For instance, Jaramillo et al. (2018) report yields of 191, 217, and 272 g/m2 in unfertilized Bahiagrass in 2014, 2015, and 2016 respectively. We currently have a system of unfertilized Bahiagrass plots outside Gainesville, Florida (northern FL), where we measured an aboveground yield of 232 g/m2 in 2019. We also note that our final harvest in this study took all the plant material to ground level and thus accounted for any earlier season production not already included in the first harvest. Overall, Bahiagrass growth is highly concentrated in the summer months, as all authors involved in this work have experienced over many years. Altogether, we disagree that we missed any substantial amount of growing season in this research, and we feel that it is therefore perfectly representative in this regard.*

R3: Why was it necessary to forgo normal analysis of variance of the dry matter yields? Exotic analyses suggest that there may have been problems with the data. The raw yield data in Table 1 suggest that there were large yield differences between the four replicates of the same plots. Is there any reason for this? Is this the reason that the authors choose to use a statistical treatment different from traditional analysis of variance that was used to analyze the data from the same plots (Vendramini et al. 2013)? The standard errors of the means in Table 1 appear to be much smaller than the crude data indicate. In the caption of Figure 3, “Where the entire 95% credible interval falls above or below zero . . .” (lines 323-4). But none of the data in Figure 3 come close to meeting this criterion, so that one can only conclude that there were no significant differences. Does this confirm my suspicion that the rather exotic statistical analysis is an attempt to disguise this conclusion? Irrespective, it is very difficult to see exactly what the results were in terms that might be useful to other readers.

*We disagree with the reviewers implied assumption that an ANOVA analysis would be a more informative way to analyze and communicate our results. The core of our analysis is just the same linear models that are omnipresent throughout applied statistics, and the reporting on coefficients (here interpreted as “effect sizes”), is canonical. For instance, figure 2 conveys all of the treatment effects for all responses of interest in one place, with points representing median estimates and the lines representing uncertainty intervals. We feel that our choice of effect coding rather than indicator coding (-0.5/0.5 versus the canonical 0/1) matches our goals for the 2X2 factorial nicely, allowing us to interpret regression coefficients for severity and frequency* ***per se****, averaging over both levels of the other variable.*

*Figure 3 is meant to be interpreted in tandem with Figure 4. In figure 4, we see which pairs of cultivars differ most strongly from each other in terms of average root production across treatments.*

*We entirely disagree with the reviewer’s suspicion that our methods were meant to “hide” or obfuscate our data. Quite the contrary, we feel that the full presentation and accounting for uncertainty (within the limits of our experimental design and data collection procedure of course) is more intellectually honest than simply sifting through ANOVA tables for p < 0.05, a procedure that has attracted considerable condemnation from leading statisticians and methodologists (see e.g. Amrhein et al. 2018). At any rate, to quell theses suspicions further, we have provided the appropriate Split-Plot ANOVA table results in a new Supplement (S2). We hope that the reviewer and other readers will be comforted that our main results (the importance of severity and cultivar) are supported by p < 0.001.*

R4: How were the plots managed for the three years from 2010 to 2013? Are these the same plots reported in Vendramini et al. 2013? If so, how were they managed in the interim?

*These are good questions. The plots are indeed the same as those reported in Vendramini et al. 2013, as we note in our materials and methods on line 182. In the interim period, the plots were mowed, weeded, and sprayed as needed around perimeter to maintain consistent cultivar composition and prevent invasion by non-target species and/or cultivars.*

R5: At over 7,300 words, the paper is far too long for the superficial nature of the work. The introduction (1232 words), materials and methods (1620 words), discussion and conclusions (1746 words) could each be reduced to about half their current length. In conclusion, the paper presents some data that may be of interest to workers in the south-eastern US. But given the short duration of the treatments, the lack of any ancillary measurements together with waterlogged soil and obscure statistical treatment, I cannot recommend that it is acceptable for publication in its present form. It may, however, merit a brief note if the authors are prepared to shorten it to about half its current length, explain the limitations of the waterlogged soil and use a more conventional statistical treatment.

*In our responses above, we have clarified our position vis a vis the utility and generalizability of our work. All in all, our findings are well within the norm for full season bahiagrass production, our environmental conditions typical of south Florida, and hence with many features in common to subtropical pasture around the world, and the main subject of our research has to do with allocation patterns and compensatory growth responses (or lack thereof), which are unlikely to only manifest in our one specific experimental area. We have restructured discussion around our hypotheses as recommended by reviewer 2, and feel that it is now easier to link the threads together.*